# Keeping the Human in the Loop: Towards Automatic Visual Monitoring in Biodiversity Research

**Joachim Denzler, Christoph Käding, Clemens-Alexander Brust**

**Computer Vision Group**
`http://www.inf-cv.uni-jena.de/`

**Michael Stifel Center Jena**
`http://www.mscj.uni-jena.de/`

Friedrich Schiller University Jena
Computer Vision Group

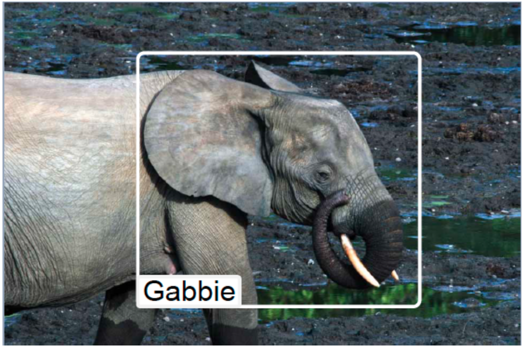
MichaelStifelCenterJena
for Data-Driven and Simulation Science

# Monitoring in Biodiversity Reserach


Classification

# Monitoring in Biodiversity Reserach


Classification




Identification

# Monitoring in Biodiversity Reserach

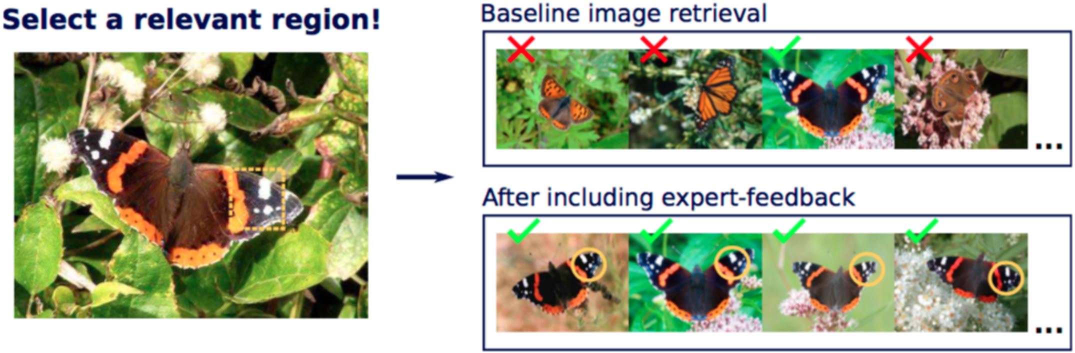

Regression . . .

# Monitoring in Biodiversity Reserach



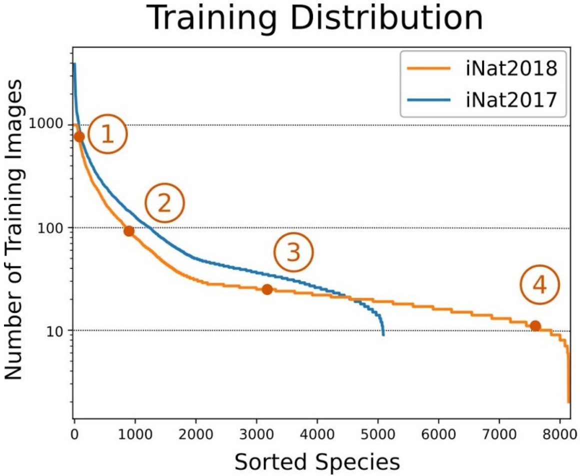

Regression . . .

# Monitoring in Biodiversity Reserach




and finally: retrieval from large data set

# Main Challenge from a Machine Learning Perspective




1 Cooper's Hawk

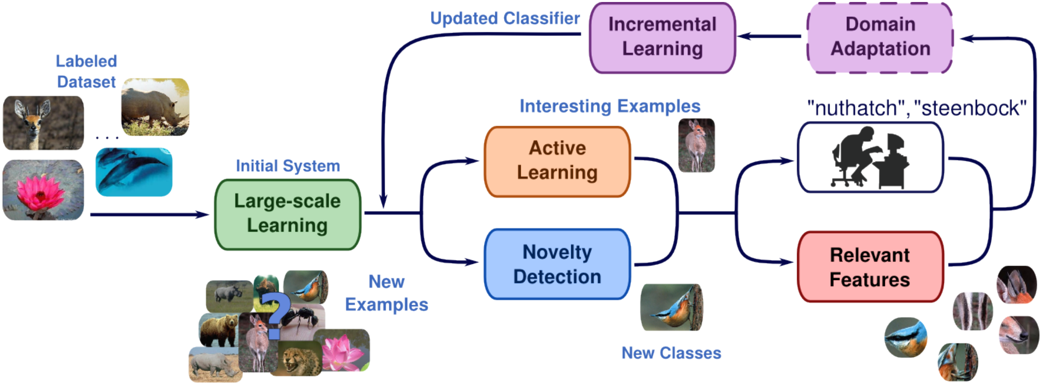
2 American Bison

3 Mallow Bindweed

4 Island Fox

from https://ai.googleblog.com/2018/03/introducing-inaturalist-2018-challenge.html

# General Theme



# Multi-class Active Learning

Idea: **Expected model output change (EMOC)**, i.e. ask for a lable for that sample that maximizes the change of the model output after knowing the label

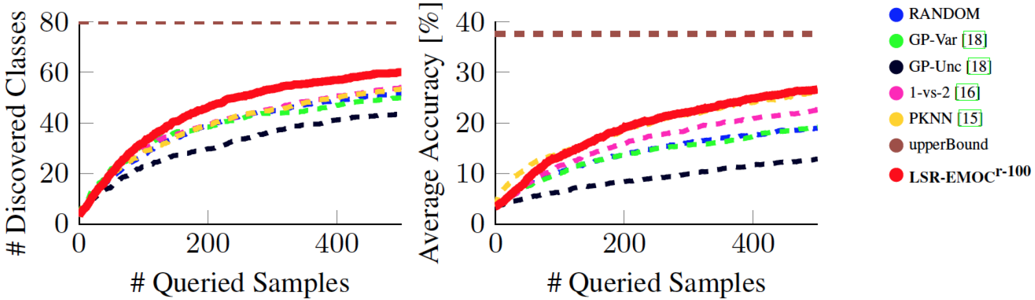
$$\triangle f(x') = \sum_{y' \in \mathbf{Y}} p(y'|f(x')) \frac{1}{|D|} \sum_{x_j \in D} L(f(x_j), f'(x_j))$$

We have sample set of labeled and unlabeled samples $D$, classifier output $f(x_j)$ given by Gaussian process classifier
We need

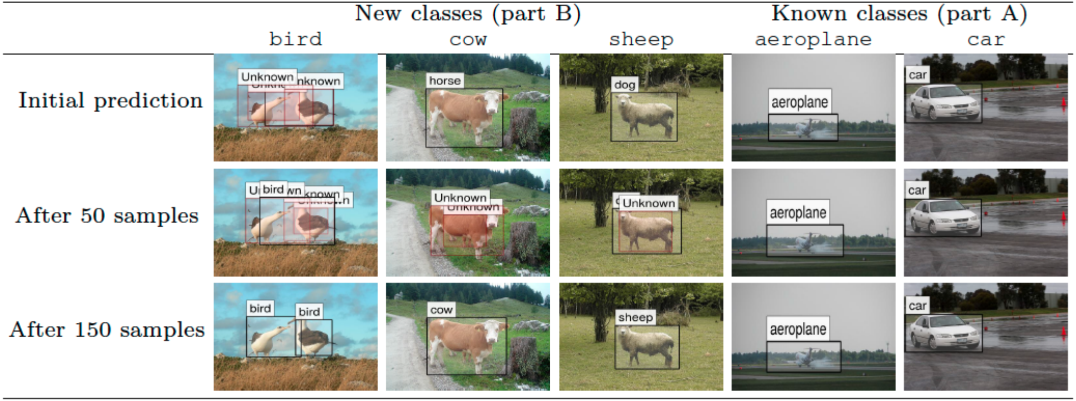
- an appropriate loss function: $L_1$-loss
- estimate for the label probability of the given sample: given by the predictive posterior of the Gaussian process classifier used for MC sampling of the labels
- efficient model update: possible in the Gaussian process framework

Käding, et al, CVPR 2015, BMVC 2018

# EMOC: Results for MS-COCO



# EMOC for Incremental Deep Detector Learning

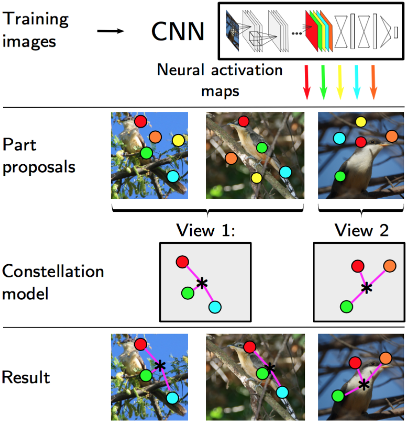



Clemens-Alexander Brust, Christoph Käding, Joachim Denzler: Active Learning for Deep Object Detection. arXiv preprint arXiv:1502.06344

# Where to Look at?



CUB200 dataset: 200 bird species, aprox. 12000 images




Simon et al. ICCV 2015, Simon et al. ICCV 2017

# Finegrained Recognition I: Birds

| Method | Training annotation Bbox | Training annotation Parts | Test annotation Bbox | Test annotation Parts | AlexNet | VGG-VG | ResNet-50 |
|---|---|---|---|---|---|---|---|
| *Previous* | | | | | | | |
| Donahue *et al.* [43], JMLR'14 | ✓ | | ✓ | | 58.8% | - | - |
| Zhang *et al.* [40], ECCV'14 | ✓ | ✓ | | | 73.9% | - | - |
| Branson *et al.* [39], BMVC'14 | ✓ | ✓ | | | 75.7% | - | - |
| Krause *et al.* [41], CVPR'15 | ✓ | | | | - | 82.0% | - |
| Zhang *et al.* [47], CVPR'16 | | | | | - | 84.5% | - |
| Liu *et al.* [50], PAMI'16 | ✓ | | ✓ | | - | 77.0% | - |
| Zhang *et al.* [73], CVPR'16 | ✓ | ✓ | ✓ | | - | 84.6% | - |
| Lin *et al.* [13], BMVC'17 | | | | | - | 85.8% | - |
| Zheng *et al.* [74], ICCV'17 | | | | | - | 86.5% | - |
| Li *et al.* [4], arXiv'17 | | | | | - | - | 86.0% |
| *Ours* | | | | | | | |
| No part modeling | | | | | 52.2% | 71.9% | 80.4% |
| Explicit (Random part selection) | | | | | 59.2% $\pm$ 0.7% | 78.6% $\pm$ 0.2% | 82.8% $\pm$ 0.3% |
| Explicit (Constellation model) | | | | | 68.5% | 81.4% | 83.4% |
| Explicit (GT parts) | | ✓ | | ✓ | 76.0% | 82.0% | 86.1% |
| Implicit ($\alpha$-pooling) | | | | | 75.1% | 86.1% | 86.5% |

Marcel Simon and Yang Gao and Trevor Darrell and Joachim Denzler and Erik Rodner. Generalized orderless pooling performs implicit salient matching. International Conference on Computer Vision (ICCV). Pages 4970-4979. 2017.

Marcel Simon, Trevor Darrell, Joachim Denzler, Erik Rodner: The whole is more than its parts? From explicit to implicit pose normalization. (submitted TPAMI).

# Finegrained Recognition II: Flowers, pets, etc.

| Method | | Oxford Flowers 102 | Oxford-IIIT Pets | Stanford 40 actions |
|---|---|---|---|---|
| *Previous* | | 84.6% [75], 86.8% [44], 91.3% [76], 94.8% [79], 96.1% [83], 96.6% [81] | 88.1% [76], 88.2% [77], 91.4% [80], 92.2% [81], | 72.0% [78], 80.9% [79], 81.7% [82] |
| *Ours* | | | | |
| AlexNet | No part modeling | 90.9% | 80.5% | 63.8% |
| | Explicit (Random part selection) | 90.4%±0.7% | 80.5%±0.7% | 63.3%±0.5% |
| | Explicit (Constellation model) | 92.1% | 82.8% | 65.6% |
| | Implicit ($\alpha$-pooling) | 94.1% | 87.0% | 68.8% |
| VGG-VD | No part modeling | 93.7% | 91.1% | 80.5% |
| | Explicit (Random part selection) | 94.6%±0.4% | 91.3%±0.3% | 82.4%±0.4% |
| | Explicit (Constellation model) | 95.5% | 91.8% | 83.3% |
| | Implicit ($\alpha$-pooling) | 97.1% | 93.2% | 86.1% |
| ResNet-50 | No part modeling | 95.7% | 93.6% | 84.1% |
| | Explicit (Random part selection) | 95.6%±0.3% | 91.1%±0.2% | 84.2%±0.4% |
| | Explicit (Constellation model) | 95.7% | 91.6% | 85.2% |
| | Implicit ($\alpha$-pooling) | **96.7%** | **94.2%** | **87.7%** |

Marcel Simon, Trevor Darrell, Joachim Denzler, Erik Rodner: The whole is more than its parts? From explicit to implicit pose normalization. (submitted to TPAMI).

# Webdemo available

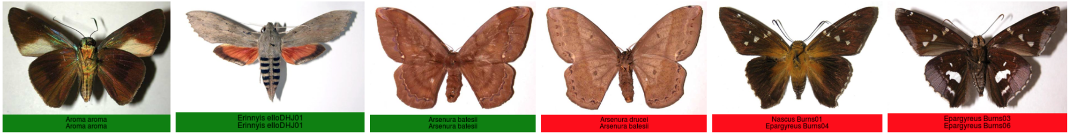


http://sigma20.inf-cv.uni-jena.de:9999/

# Finegrained Recognition III: Moths classification

| dataset | #classes | #images | #images for training | accuracy (global) | accuracy (pyramid) |
|---|---|---|---|---|---|
| Ecuador moth dataset [3] | 675 | 2120 | 1445 | 55.7% | 53.5% |
| Costa rica dataset [5] | 331 | 3224 | 992 | 79.5% | 82.1% |

Table 1: Categorization results for the two biodiversity datasets (butterflies and moths) of [3] and [5].



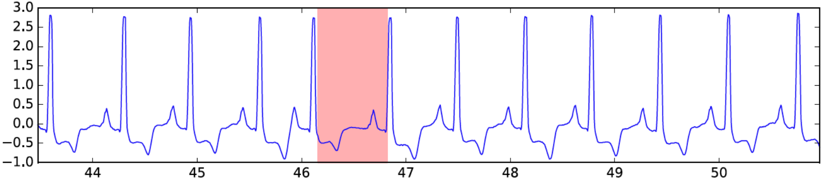

Rodner, Simon, Brehm, Pietsch, Wägele, Denzler. Fine-grained Recognition Datasets for Biodiversity Analysis CVPR Workshop on Fine-grained Visual Classification (CVPR-WS). 2015.

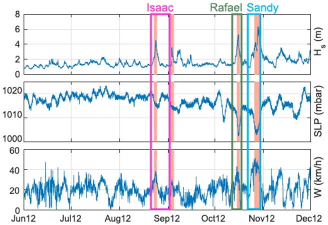
# Explaining the Results ...

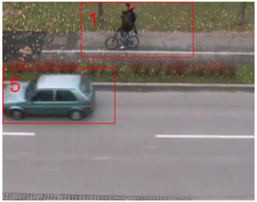

# Semantic Image Retrieval

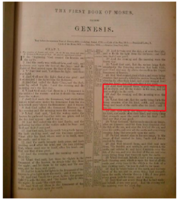

Barz, Denzler: *Hierarchy-based Image Embeddings for Semantic Image Retrieval*. arXiv preprint arXiv:1502.06344

# Detection of Anomalies




Barz, Rodner, Guanche Garcia, Denzler: *Detecting Regions of Maximal Divergence for Spatio-Temporal Anomaly Detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

# Monitoring: Herbivorous Mammals

**Study design**

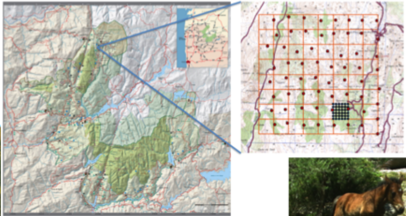
100 cameras in nested grid (3.5 x 3.5 km)

Evenly distributed over different habitat types (grass land, heath, forest)

More than 8,000 camera days
(individual cameras working between 4 and 125 days from April to September)

Total of 412,217 images recorded

Page 2



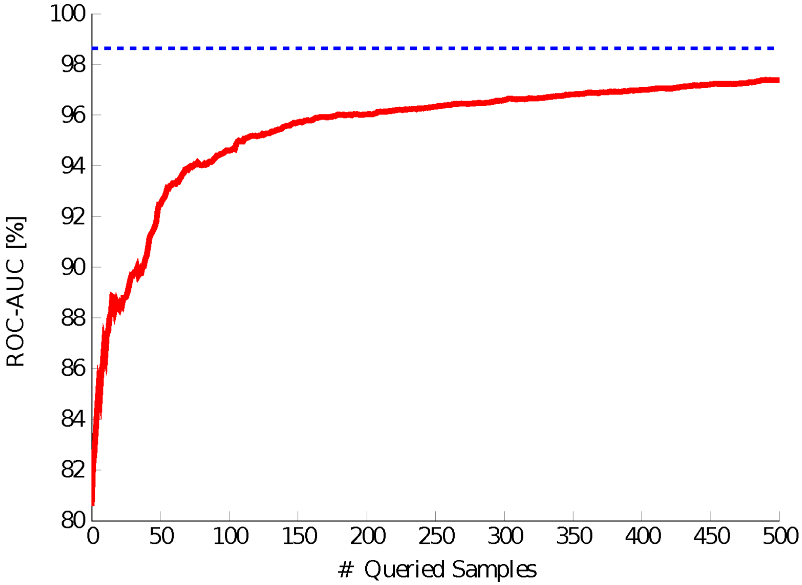
http://www.parks.it/world/PT/pn.PenedaGeres/mapl.jpg

Joint work with mit Andrea Perino, IDiv (thanks for the slide!)

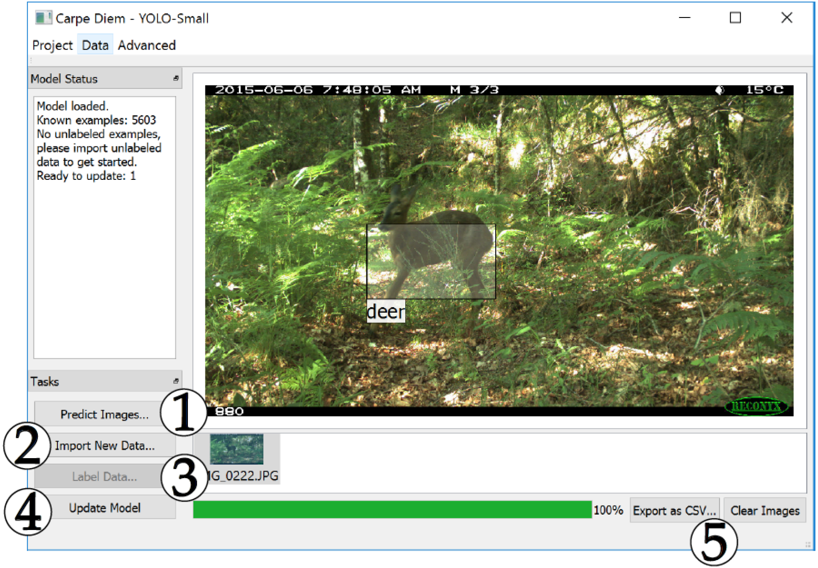
# Is there an animal in the image?


Christoph Käding, Alexander Freytag, Erik Rodner, Andrea Perino, Joachim Denzler: Large-scale Active Learning with Approximations of Expected Model Output Changes. German Conference on Pattern Recognition (GCPR). 2016.

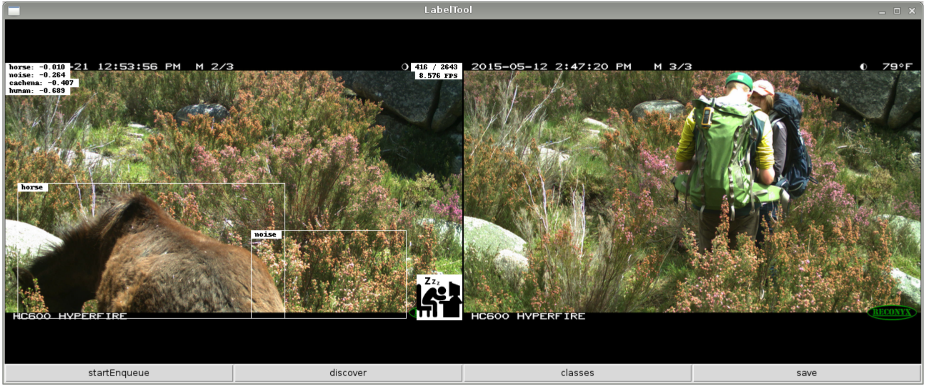
# Is there an animal in the image?



# Where is the animal in the image?



# Monitoring Animals in the Wild




C. Käding, A. Freytag, E. Rodner, A. Perino, J. Denzler. Large-scale Active Learning with Approximated Expected Model Output Changes. German Conference on Pattern Recognition (GCPR). Pages 179-191. 2016.

# Monitoring Animals in the Wild


C. Käding, A. Freytag, E. Rodner, A. Perino, J. Denzler. Large-scale Active Learning with Approximated Expected Model Output Changes. German Conference on Pattern Recognition (GCPR). Pages 179-191. 2016.

# Summary



- Elements for (semi-)automatic monitoring are already available today
- Keeping the human in the loop is important in many applications
- Methods have to be generic and able to live in an open world

# Summary

- Elements for (semi-)automatic monitoring are already available today
- Keeping the human in the loop is important in many applications
- Methods have to be generic and able to live in an open world
- Systems already deployed in various areas
  - Monitoring of herbivorous mammals (Idiv, Leipzig)
  - Identification and age estimation of gorillas (WCS, Mbeli Bai Study)
  - **Identification of elefants** (Cornell Lab of Ornithology, Elephant Listening Project)

# Summary

- Elements for (semi-)automatic monitoring are already available today
- Keeping the human in the loop is important in many applications
- Methods have to be generic and able to live in an open world
- Systems already deployed in various areas
  - Monitoring of herbivorous mammals (Idiv, Leipzig)
  - Identification and age estimation of gorillas (WCS, Mbeli Bai Study)
  - **Identification of elefants** (Cornell Lab of Ornithology, Elephant Listening Project)
- Demos and code available for most parts of research
  - Bird classification: `http://sigma20.inf-cv.uni-jena.de:9999/`
  - Interactive annotation: `http://triton.inf-cv.uni-jena.de/LifelongLearning/carpediem`
  - Interactive image retrieval: `http://sigma24.inf-cv.uni-jena.de:8080/`
  - Anomaly detection: `https://github.com/cvjena/libmaxdiv`

# People

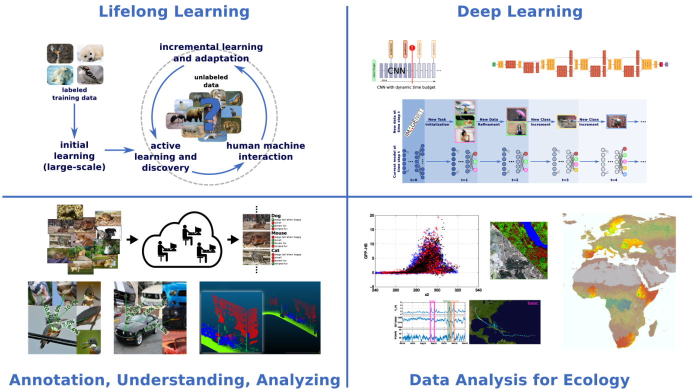
# Further Information

For more information (details, datasets, papers, videos) please visit

1. Computer Vision Group: http://www.inf-cv.uni-jena.de
2. Youtube Channel CVG: https://www.youtube.com/channel/UCpnLVdxmvF0zEHHIVfqSxag
3. Google+: https://plus.google.com/+ComputerVisionGroupJena/videos
4. Michael Stifel Center Jena: http://www.mscj.uni-jena.de

Projects are funded by EU H2020, DFG, BMBF, BMWi, TAB

# Thank you for your attention!




http://www.inf-cv.uni-jena.de
https://plus.google.com/+computervisiongroupjena